

最大熵谱的优良特性和预报误差过滤 系数阶数的确定*

缪 锦 海

(北京大学地球物理系)

提 要

最大熵谱分析比通常谱分析分辨率高, 峰值偏离小, 特别适宜于短资料或长周期的谱分析, 这对气候分析和中长期过程的周期分析是有显著意义的。

最大熵谱分析的预报误差过滤系数的阶数, 可以在最终预报误差 $(FPE)_m$ 的前后两个极小值的谱估计一致的前提下, 用前一个极小值的 m 数来确定。

一、引 言

谱分析在气候的周期性分析和气象场统计特性的分析中, 都是重要的统计方法。目前, 在气象上常用的谱分析方法, 是通过资料的自相关函数乘以某些窗函数, 然后经过富氏变换而得到功率谱估计的。这种由 Blackman 和 Tukey^[1](1959) 提出的谱分析方法, 实际上对资料作了这样的假定: 在所得资料之外认为是周期的或是零的资料。这样的假定不利于估计真实谱, 不利于谱的分辨。1967 年 Burg^[2]提出了最大熵方法(Maximum Entropy Method 简称 MEM)。熵是对于随机过程的单位时间的平均信息。最大熵方法没有上述的假定, 而是根据现有资料, 使得平稳过程的熵最大推得无限个滞后的自相关函数, 从而得到谱的估计。Parzen^[3](1968)从另一观点即最小二乘预报误差过滤的自回归谱估计得到相同的结果。这实际上改变了通常谱分析方法的固定窗即不随着资料的真实谱的特性而变化的缺点。当估计某一频率的功率时, 调整窗的特性使得其它频率功率影响最小, 这样就有利于得到真实谱。因而, 具有比通常的谱方法显著的分辨能力, 对于资料长度比较短的情况尤为明显。正因为有这样突出的长处, 最大熵谱分析受到普遍的重视。在有些学科的应用上已取得显著的成功。Ulrych 和 Bishop^[4](1975)作了全面的评论。Hayashi(林 良一)^[5](1977)把最大熵谱应用到气象方面, 进行空间-时间的功率谱分析, 取得了好的结果。最近我国曹鸿兴、罗乔林讨论了气象资料的最大熵谱分析(在内部会议上)。

本文在简要说明最大熵谱方法的基础上, 利用理想的谱和实际例子, 进一步具体讨论最大熵谱方法的优良特性和局限性, 并对计算最大熵谱的重要问题即确定预报误差过滤系数的阶数, 根据试验提出新的意见。

* 本文于 1978 年 5 月 24 日收到。

二、最大熵谱分析

1. 最大熵谱

对于平稳过程的熵为

$$H = \frac{1}{4f_N} \int_{-f_N}^{f_N} \ln S(f) df \quad (1)$$

这里 f_N 是 Nyquist 频率, $f_N = 1/(2\Delta t)$, Δt 是样本间隔, f 为频率, $S(f)$ 为功率谱。

对于平稳过程的功率谱和自相关函数的关系为

$$S(f) = \int_{-\infty}^{\infty} \rho(\tau) \exp(-i 2\pi f\tau) d\tau \quad (2)$$

$$\rho(\tau) = \int_{-f_N}^{f_N} S(f) \exp(i 2\pi f\tau) df. \quad (3)$$

在已知 m 个滞后的自相关函数 ($\rho(0), \rho(1), \dots, \rho(m-1)$) 的约束条件下, 使得平稳过程的熵最大的功率谱称为最大熵谱。对(1)式在 m 个(3)式的约束条件下, 按照变分原理, 可求得最大熵谱(推导详见 Smylie 等^[6]或 Edward 和 Fitelson^[7])

$$S(f) = \frac{p_m \Delta t}{\left| 1 + \sum_{j=1}^m \gamma_j \exp(-i 2\pi f j \Delta t) \right|^2}, \quad (4)$$

其中 p_m 和 $\gamma_j (j=1, \dots, m)$ 由(5)式决定

$$\begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_m \\ \rho_1 & \rho_0 & \cdots & \rho_{m-1} \\ \vdots & \vdots & & \vdots \\ \rho_m & \rho_{m-1} & \cdots & \rho_0 \end{pmatrix} \begin{pmatrix} 1 \\ \gamma_1 \\ \vdots \\ \gamma_m \end{pmatrix} = \begin{pmatrix} P_m \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (5)$$

在(5)式中可以分为

$$P_m = \rho_0 + \gamma_1 \rho_1 + \cdots + \gamma_m \rho_m \quad (6)$$

和

$$\begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_{m-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{m-2} \\ \vdots & \vdots & & \vdots \\ \rho_{m-1} & \rho_{m-2} & \cdots & \rho_0 \end{pmatrix} \begin{pmatrix} -\gamma_1 \\ -\gamma_2 \\ \vdots \\ -\gamma_m \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_m \end{pmatrix}. \quad (7)$$

这(7)式与 m 阶自回归模式

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_m x_{t-m} + e_m \quad (8)$$

的系数 (a_1, a_2, \dots, a_m) 满足的方程

$$\begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_{m-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{m-2} \\ \vdots & \vdots & & \vdots \\ \rho_{m-1} & \rho_{m-2} & \cdots & \rho_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_m \end{pmatrix} \quad (9)$$

是相同的。因而

$$\gamma_1 = -a_1, \gamma_2 = -a_2, \dots, \gamma_m = -a_m.$$

(6)式可以改写为

$$p_m = \rho_0 - (a_0\rho_1 + \dots + a_m\rho_m) \quad (10)$$

所以 p_m 是 $m+1$ 长的预报误差过滤器 $(1, -a_1, \dots, -a_m)$ 的输出功率, 简称为预报误差功率。

从另一观点来看, 由于这功率谱是通过过滤器 $(1, -a_1, \dots, -a_m)$ 得到, 而这过滤器是由自回归模式的最小二乘预报误差的条件下产生的, 这过滤器的系数称为最小二乘预报误差过滤系数或预报误差过滤系数。

2. 预报误差过滤系数的估计

求 $\gamma_1, \dots, \gamma_m$ 的方法, 可以用 Yule-Walker 估计的方法(Y-W 方法)。先计算 $\rho(\tau)$ 的估计值 $\hat{\rho}(\tau)$

$$\hat{\rho}(\tau) = \frac{1}{N-\tau} \sum_{t=1}^{N-\tau} x_{t+\tau} x_t \quad (11)$$

其中 x_t 是标准化序列, 然后把 $\hat{\rho}(\tau)$ 代入方程(9), 可求得 (a_1, \dots, a_m) 。

Burg^[8](1968)提出另一种新的估计方法。对于最大熵谱满足的(5)式, 当在第 $m-1$ 步求解后, 已知 $\rho_0, \rho_1, \dots, \rho_{m-1}$, 而在下一步即第 m 步时, 就包含有 $(a_1, a_2, \dots, a_m, \rho_m$ 和 $p_m)$ $m+2$ 个未知数。而(5)式只有 $m+1$ 个方程, 所以要求增加一个附加判据才能得到唯一解。Burg 提出使总的预报误差过滤的平均功率最小作为附加判据。

对于一阶向前预报误差平均功率

$$\frac{1}{N-1} \sum_{t=1}^{N-1} (x_{t+1} - a_{11}x_t)^2.$$

一阶向后预报误差平均功率

$$\frac{1}{N-1} \sum_{t=1}^{N-1} (x_t - a_{11}x_{t+1})^2.$$

总的一阶预报误差过滤的平均功率为

$$\pi_1 = \frac{1}{2(N-1)} \sum_{t=1}^{N-1} [(x_t - a_{11}x_{t+1})^2 + (x_{t+1} - a_{11}x_t)^2].$$

使 π_1 达到最小而求 a_{11} , 即

$$\frac{\partial \pi_1}{\partial a_{11}} = 0$$

可得

$$a_{11} = \frac{2 \sum_{t=1}^{N-1} x_t x_{t+1}}{\sum_{t=1}^{N-1} (x_t^2 + x_{t+1}^2)}. \quad (12)$$

对于 m 阶预报误差过滤的平均功率为

$$\pi_m = \frac{1}{2(N-m)} \sum_{t=1}^{N-m} \left[\left(x_t - \sum_{k=1}^m a_{mk} x_{t+k} \right)^2 + \left(x_{t+m} - \sum_{k=1}^m a_{mk} x_{t+m-k} \right)^2 \right] \quad (13)$$

根据最小化条件

$$\frac{\partial \pi_m}{\partial a_{mm}} = 0$$

可得

$$a_{mm} = \frac{2 \sum_{t=1}^{N-m} b_{mt} b'_{mt}}{\sum_{t=1}^{N-m} (b_{mt}^2 + b'_{mt}{}^2)}, \quad (14)$$

其中

$$b_{mt} = \sum_{k=0}^m a_{m-1k} x_{t+k} = \sum_{k=0}^n a_{m-1m-k} x_{t+m-k} \quad (15)$$

$$b'_{mt} = \sum_{k=0}^m a_{m-1k} x_{t+m-k} = \sum_{k=0}^m a_{m-1m-k} x_{t+k}$$

而

$$a_{mk} = a_{m-1k} - a_{mm} a_{m-1m-k}, \quad k=1, 2, \dots, m=1 \quad (16)$$

根据(15), (16)式可以得到递推公式

$$b_{mt} = b_{m-1t} - a_{m-1m-1} b'_{m-1t} \quad (17)$$

$$b'_{mt} = b'_{m-1t+1} - a_{m-1m-1} b_{m-1t+1}$$

初始情况为

$$b_{0t} = b'_{0t} = x_t \quad t=1, 2, \dots, N$$

$$b_{1t} = x_t \quad (18)$$

$$b'_{1t} = x_{t+1}$$

由(16)式代入(5)可得 p_m 的递推公式

$$p_m = p_{m-1} (1 - a_{mm}^2) \quad (19)$$

$$p_0 = \frac{1}{N} \sum_{t=1}^N x_t^2$$

Andersen^[9](1974)作了详细的推导。

3. 自回归阶数即预报误差过滤系数的阶数

Ulrych等^[4]提出用Akaike^[10](1969)的最终预报误差(Final Prediction Error FPE)的定阶方法来确定最大熵谱方法的过滤系数的阶数。最终预报误差

$$(\text{FPE})_m = \frac{N+m}{N-m} p_m \quad (20)$$

取 $(\text{FPE})_m$ 的第一个极小值或最小值时的 m 数作为过滤系数的阶数。

三、最大熵谱的优良特性

我们通过理想的正弦波加上白噪声组成的时间序列和一些实际的时间序列来对比最

大熵谱和通常的谱分析，从中可以看到最大熵谱有以下的显著特性和局限性：

1. 最大熵谱比通常的谱分析具有突出的分辨能力。如图 1，为频率 $f_1=0.04$ 和频率 $f_2=0.06$ 的相同振幅的两个正弦波和标准差为两个正弦波的 $\frac{1}{9}$ 的白噪声组成的时间序列(资料数 N 为 50)的最大熵谱和通常谱。最大熵谱 Burg 法的功率谱能正确又清楚地分辨出有频率 $f_1=0.04$ 和频率 $f_2=0.06$ 的两个峰值。而通常谱分析只能给出 $f_1=0.04$ 的一个峰值。最大熵谱甚至能分辨频率 $f_1=0.04$ 和频率 $f_2=0.05$ 的两个非常接近的正弦波(资料数为 50)，见图 2。

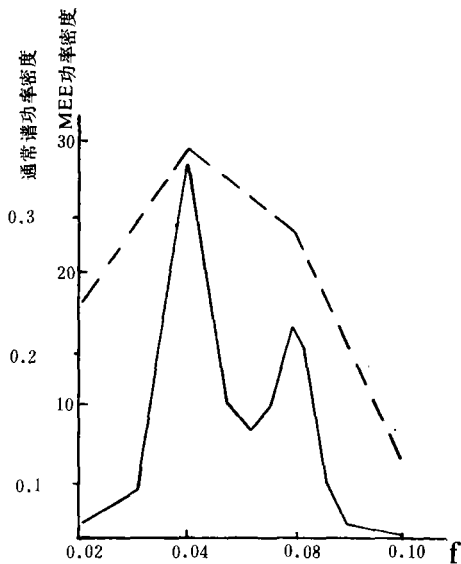


图 1 频率 $f_1=0.04$ 和频率 $f_2=0.06$ 的正弦波加上白噪声的功率谱 ($N=50$)
——MEM
-----通常谱

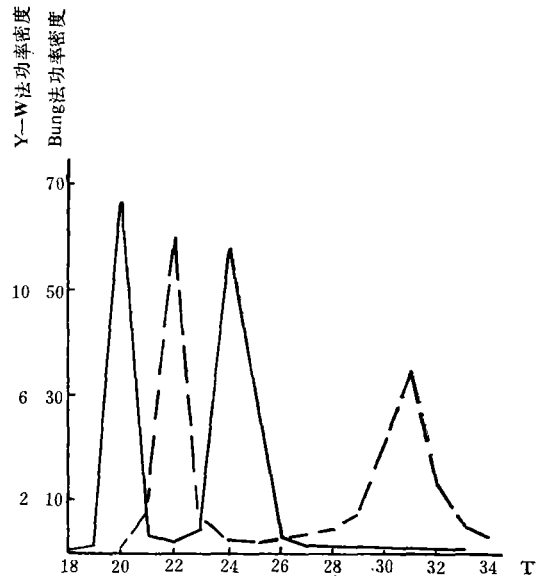


图 2 频率 0.04 和频率 0.05 的正弦波加上白噪声的 MEM 功率谱 ($N=50$)
——Burg 法
-----Y-W 法

分辨能力与资料长度有关。图 3 为频率 0.04 和频率 0.08 加上白噪声的通常谱。在资料长度 $N=50$ 时只能初略的分辨出不明显的两个峰值， $N=100$ 时才能分辨清楚有两个峰值。显然，在 $N=25$ 时是不能分辨的。对通常谱分析的分辨能力的试验情况见表 1， \times 表示 f_1 与 f_2 不能分辨， \surd 表示初略的分辨， \checkmark 表示能明显分辨。这样，总结通常谱分析(经过光滑的)的分辨能力为 $f_1 - f_2 = 2/N$ ，才能初略的分辨。所以，最大熵谱的分辨能力比通常的谱分析(经过光滑的)至少要大 4 倍。

2. 最大熵谱 Burg 法的谱估计比 Y-W 法和通常谱估计更符合真实的谱特征，谱的峰值的偏离小。从图 2 中可以看到 Y-W 法的峰值有明显的偏离现象， $f_1=0.04$ (周期 $T=20$) 偏到 $T=22$ 处， $f_2=0.05$ ($T=25$) 偏到 $T=32$ 处。通常谱分析的偏离也比较显著，正如 Toman^[11](1965)指出，在资料数小于正弦波周期长度的 0.58 倍时，最大谱值将出现在零频率处。这与我们的试验是一致的。

最大熵谱 Burg 法在资料长度为正弦波周期长度 $\frac{1}{4}$ — $\frac{1}{2}$ 时也会出现偏离现象。如图

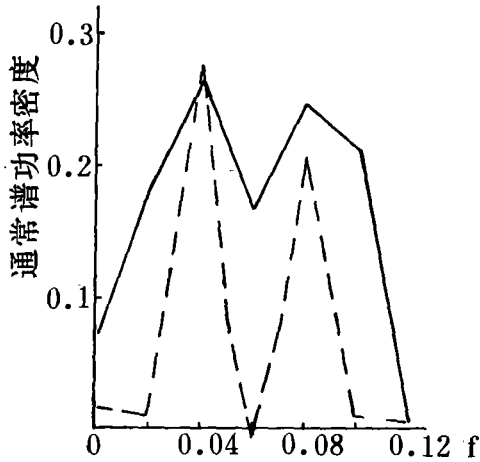


图3 频率0.04和频率0.08加上白噪声的通常谱
—— $N=50$ 的通常谱
----- $N=100$ 的通常谱

表1 通常谱的分辨能力与资料长度
 N 的关系。
(说明见文内)

$f_1 - f_2$	N	50	100
$0.05 - 0.04 = 0.01$		×	×
$0.06 - 0.04 = 0.02$		×	✓
$0.07 - 0.04 = 0.03$			✓
$0.08 - 0.04 = 0.04$		✓	✓

4 为 $f_1 = 0.01$ 和 $f_2 = 0.1$ 加上白噪声的资料长度为 50 的功率谱。其中 $f_1 = 0.01$ ($T = 100$ 比资料数长一倍) 的峰值出现在 0.009 处。我们计算了 $T = 100$ 、 $N = 25$ 的初位相为零的正弦波的最大熵谱, 其峰值出现在 $T = 95$ 处。在我们的试验中, 当资料长度为正弦波周期长度的 $\frac{1}{3} - \frac{1}{2}$ 时, 偏离程度约为频率的 5—10% 左右。这与 Chen 和 Stegen^[12] (1974) 的试验是一致的。他们并指出在资料长度为周期的 $\frac{1}{4}$ 时, 偏离程度与初位相的关系很大, 有时偏离可达频率的一倍左右。

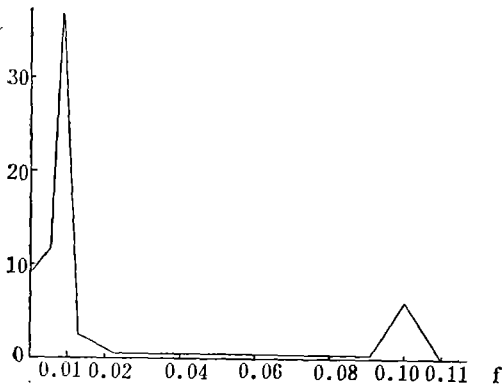


图4 频率 $f_1 = 0.01$ 和频率 0.1 的正弦波加上白噪声的最大熵谱 ($N = 50$)

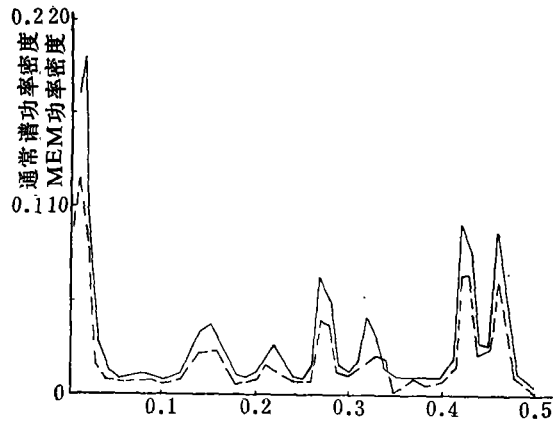


图5 一百年的在 130°E 、 20°N 处的地面气压的谱估计

——最大熵谱
-----通常谱

3. 对于短资料 and 低频部份的功率谱的分析, 最大熵谱 (Burg 方法) 具有独特的分析能力。在图 4 中, 资料长度为长周期的一半, 并混在高频的波动中, 也能得到满意

的功率谱估计。这对气候和中长期预报的周期分析是有意义的。

图 5 是在 20°N, 130°E 处的一百年地面气压的谱估计。最大熵谱分析能估计出有 75 年左右的长周期振动, 而通常谱分析只能在周期 100 年($f=0.01$)出现峰值, 估计不清楚。类似地, 我们计算了上海一百年降水量(等级)的谱估计, 最大熵谱能估计出 31 年左右的周期, 通常谱分析只能粗糙地估计为 30—40 年周期。从图 5 可以看到, 两种谱在其他部份是相似的, 但最大熵谱要明显些。

四、预报误差过滤系数阶数的确定

最大熵谱通过根据原始的资料特征进行外推来克服通常谱认为原资料之外是周期的或零的特征的弱点。它的外推方法是用最小二乘预报误差过滤的方法即自回归模式估计的方法。这样, 过滤系数阶数的不同, 外推的结果就不同, 因而谱估计就不同。所以, 过滤系数阶数的确定是值得重视的问题, 它影响着最大熵谱的实际运用。Chen 和

Sthegen^[12](1974)提出阶数在 6—15 之间比较适当的意见。Ulyrch 和 Bishop^[4](1975)提出用 $(FPE)_m$ 的第一个极小值或最小值时的 m 数作为阶数的意见。但是, 在我们分析的不同频率、不同资料长度的理想例子和实例中, 发现他们的方法并不能都得到正确的谱估计。图 6 是不同频率、不同资料长度的各个阶数 m 时的功率谱峰值的频率的变化图。图中(1), (2), ……表示 FPE 的第一个极小值, 第二个极小值……。图中[]表示 FPE 的最小值。从图中可以看到有四个例子的阶数大于 15 才能得到正确的谱估计。在图 6 中, 只有图 6A 这一个例子($f_1=0.1, f_2=0.2$)的 FPE 的第一个极小值的阶数时谱估计是正确的谱估计。其它的例子, 用第一个极小值的阶数来估计都不能得到正确的谱估计。图 6B, 图 6C 是第二个极小值情况是正确的谱估计(另有三例 $f_1=0.33, f_2=0.38; f_1=0.04, f_2=0.06; f_1=0.01, f_2=0.1$, 图略)。其它例子都是第三个极小值或多于第三个极小值的阶数时才是正确的谱估计。从图 6E, 图 6F 两个实例来看, 采用 FPE 的最小值的阶数(分别为 7 与 2)都不能得到正确的谱估计。

虽然, Ulyrch 等提出的方法是比较合理的, 因为在 FPE 为极小值或最小值时的预报误差功率(即方差)为极小, 这样, 根据原来的资料特征的外推就可能比任意规定的阶数要合理些。但

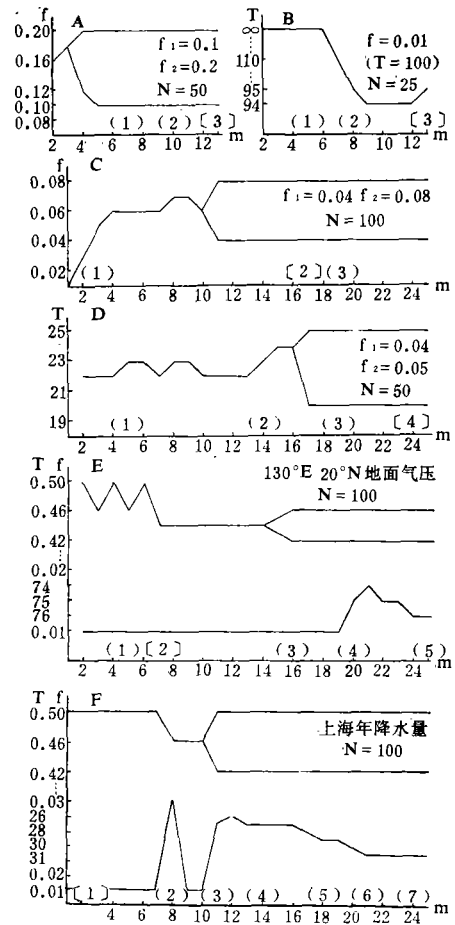


图 6 不同频率、不同资料长度的功率谱峰值的频率随阶数 m 的变化图。(1), (2)……表示 FPE 的第一个极小值, 第二个极小值……, [] 表示 FPE 的最小值。

是,通过以上例子的试验说明这方法在有些情况谱估计不正确。如何改进 Ulyrch 等提出的方法呢? 从我们试验的所有例子中, 可以看到一个共同的特点: 当 FPE 的前后两个极小值的谱估计一致时, 这样的谱估计是合适的。如图 6A 中第一个极小值与第二个极小值的谱估计是一致的, 图 6B, 图 6C 中第二个极小值和第三个极小值的谱估计是一致的等等, 这时的谱估计是正确的。另外, 常常是前一个极小值的谱估计的数值比较合适。这说明不仅要预报误差功率极小, 而且还要这些功率极小时的谱估计是稳定的。这样, 才能更好地反映外推的可靠性, 充分地反映最大熵谱分析的优点, 得到比较正确的谱估计。因此, 我们提出在 $(FPE)_m$ 的前后两个极小值的谱估计一致时, 用前一个极小值的 m 数定为预报误差过滤系数的阶数的方法。

五、结 语

最大熵谱分析比通常谱分析分辨率高, 峰值偏离小, 特别适宜于进行短资料的谱分析。最大熵谱的分辨率至少是通常谱分析的 4 倍, 在资料长度为周期长度的 $\frac{1}{4}$ — $\frac{1}{2}$ 时, 峰值偏离约为频率的 5%—10% 左右。

最大熵谱分析的预报过滤系数的阶数, 可在 FPE 的前后两个极小值的谱估计一致条件下, 用前一个极小值的 m 数来确定。

致谢: 对王绍武、黄嘉佑、赵宗慈等同志给予的多方面的帮助和支持, 谨致谢意。

参 考 文 献

- [1] R. B. Blackman and J.W. Tukey, The measurement of power spectra from the point of view of communication engineering, New York, Dover, 1959.
- [2] J.P. Burg, Maximum entropy spectral analysis, Paper presented at the 37th Annual International Meeting, Soc. of Explor. Geophys. Oklahoma City, Okla., Oct. 31, 1967.
- [3] E. Parzen, Multiple time series modeling, Tech. Rep. 12 on Contract Nonr-225-(80), Stanford University, 1968.
- [4] T.J., Ulyrch and T.N., Bishop, Maximum entropy spectral analysis and autoregressive decomposition, *Rev. of Geophys. and Space Phys.* 13 (1975) 183—200.
- [5] Y., Hayashi (林 良一), Space-time power spectral analysis using the maximum entropy method, *J. Meteor. Soc. Japan*, 55 (1977) 415—420.
- [6] D.E., Smylie, G. K. C. Clarke, and T. J., Ulyrch, Analysis of irregularities in the earth's rotation, in *Methods in Computational Physics*, 13 (1973), 391—430, Academic, New York.
- [7] J. A., Edward, and M. M. Fitelson, 1973: Notes on maximum-entropy processing, *IEEE Trans. Inform. Theory*, IT-19, 232.
- [8] T. P., Burg, A new analysis technique for time series data, Paper presented at Advanced Study Institute on Signal Processing, NATO, Eindhoven, Netherlands, 1968.
- [9] M. Andersen, On the calculation of filter coefficient for maximum entropy analysis, *Geophysics*, 39(1974), 69—72.
- [10] H., Akaike, Fitting autoregressive models for prediction, *Ann. Inst. Statist. Math.* 21(1969), 243—247.
- [11] K., Toman, Spectral shifts of truncated sinusoids, *J. Geophys. Res.*, 70(1965), 1749—1750.
- [12] W. Y., Chen, and G. R. Stegen, Experiments with maximum entropy power spectra of sinusoids, *J. Geophys. Res.*, 79(1974), 3019—3022.

THE ADVANTAGE OF MAXIMUM ENTROPY METHOD AND DETERMINATION OF THE ORDER OF ITS PREDICTION ERROR COEFFICIENT

Miao Jin-hai

(Department of Geophysics, Beijing University)

Abstract

The advantage of maximum entropy method (MEM) of spectral analysis is that it gives finer resolutions and less peak discrepancy than the conventional spectral analysis. MEM is particularly useful for short data lengths and long period spectral analysis. It bears special significance for periodic analysis of climate and medium-range and long-range process.

The order of the coefficient of MEM prediction error may be determined by the order of the former minimum under the condition that the power of the former minimal final prediction error of spectral estimate is consistent with the latter.